	NUR 1 UIII 1	MARE 5	WHE 8	WHE7	 Project 3 anomali

Project Roadmap: Multi-Class Prediction of Obesity Risk

The Team:

- 1. Product Manager Aasna
- 2. Machine Learning Engineer Arham
- 3. ML Ops Krishan
- 4. Data Engineer Yash
- 5. Cloud SME Nandani
- 6. Business Analyst Mahrukh

Objective: Develop a predictive analytics solution that can classify individuals into various risk categories for obesity based on their health data. This solution aims to aid healthcare providers in identifying highrisk individuals and tailoring preventive measures accordingly.

Scope

Data Collection and Preparation: Utilize datasets comprising health metrics to train predictive models.

Model Development and Experimentation: Apply machine learning techniques to develop and refine prediction models.

Testing and Validation: Evaluate models in a simulated realworld environment.

Deployment: Implement models both locally and on the cloud for accessibility and scalability.

User Interaction: Develop a frontend application for easy access to prediction results.

Technologies Used

- 1. Python, Jupyter Notebooks for data analysis and model training.
- 2. MLFlow for experiment tracking.
- 3. PyCaret, LightGBM, XGBoost, CatBoost for model building and training.
- 4. Optuna for hyperparameter optimization.
- 5. Docker for containerization of the application.
- 6. Kafka for simulating realtime data streaming.
- 7. Git, GitHub Actions for version control and CI/CD
- 8. Azure Machine Learning, Azure Blob Storage for cloud deployment.
- 9. Streamlit for developing the user interface.

Description of Phases:

Phase 1: Planning and Design (Week 1-2)

During this initial phase, the team establishes the foundation of the project. The Product Manager sets the project's vision and milestones. The Machine Learning Engineer and Business Analyst research technical feasibility and market requirements, respectively. The Data Engineer and Cloud SME lay the groundwork for data handling and cloud infrastructure, ensuring all systems align with the project's technical needs.

Phase 2: Data Preparation and Infrastructure Setup (Week 2-4)

In this phase, the team focuses on setting up the necessary infrastructure and preparing the data for analysis. The Data Engineer builds data ingestion pipelines, while the Cloud SME ensures the cloud setup is optimized for scalability and security. The team begins collecting and preprocessing data to prepare for model development.

Phase 3: Feature Engineering and Model Prototyping (Week 5)

Feature engineering and initial model prototyping are conducted. The Machine Learning Engineer explores and selects features that will effectively predict obesity risk, while developing initial models to test their efficacy. The ML Ops specialist begins to establish systems for tracking and managing these models.

Phase 4: Model Refinement and Experimentation (Week 6)

This phase is critical for refining the models through extensive experimentation and tuning. The team iterates on models, optimizing their performance through advanced analytical techniques and continuous testing. The ML Ops focuses on solidifying the model deployment pipeline and version control systems.

Phase 5: Deployment Preparation and Testing (Week 7)

Preparation for deployment involves finalizing the model, setting up continuous integration/continuous deployment (CI/CD) pipelines, and ensuring all systems are robust and secure. The team conducts final stress tests to ensure the infrastructure is ready for a smooth transition to production.

Phase 6: Model Deployment and Monitoring (Week 8)

The model is deployed to a production environment. This phase includes rigorous monitoring of the model's performance and quick resolution of any issues. The team focuses on ensuring the model operates efficiently and effectively.

Tasks:

Phase 1: Planning and Design

Aasna (Product Manager): Develop and document the project's vision and mission statement, articulate the value proposition of the obesity risk prediction product, and establish the overall product strategy and milestones.

Arham (Machine Learning Engineer): Conduct a technical feasibility study on multiclass classification techniques and their applicability to obesity risk prediction, and document the requirements for initial data sets.

Krishan (ML Ops): Draft a preliminary machine learning pipeline, including data ingestion, model training, evaluation, and deployment processes, and select appropriate tools and platforms for ML Ops.

Yash (Data Engineer): Prepare a detailed plan for the data architecture, including sourcing, storage, and flow of data, and create a schema for the data to be used in model training.

Nandani (Cloud SME): Define the cloud resource requirements for the project, including compute, storage, and networking, and draft a plan for cloud security and compliance.

Mahrukh (Business Analyst): Analyze industry trends in health tech and obesity prediction, identify key market segments and user personas, and draft a report on the competitive landscape.

Phase 2: Data Preparation and Infrastructure Setup

Aasna: Develop a comprehensive project management dashboard to track progress across teams, outline the project risk management plan, and initiate weekly crossfunctional meetings.

Arham: Select and prepare datasets for initial exploration, identify potential data quality issues, and create a preprocessing plan to handle missing data and outliers.

Krishan: Establish a version control system for data and models, and set up the initial machine learning environment in collaboration with Yash and Nandani.

Yash: Implement automated data ingestion pipelines, ensuring data integrity and enabling initial datasets to be populated and ready for exploration.

Nandani: Deploy the foundational cloud infrastructure, ensuring that it is scalable, secure, and meets the requirements set by Arham and Krishan for model training and data processing.

Mahrukh: Begin a deep dive into the available data, identifying key features that could influence obesity risk prediction and providing a detailed report on initial findings to the ML team.

Phase 3: Feature Engineering and Model Prototyping

Aasna: Oversee the prototype development, maintain the product backlog, and ensure that the feature engineering is aligned with the user needs identified by Mahrukh.

Arham: Apply various feature selection and feature extraction techniques, document their impact on model performance, and build a series of prototype models using different algorithms.

Krishan: Configure an experiment tracking system to log model versions, parameters, and performance metrics, and create a feedback loop for continuous improvement.

Yash: Refine the data pipelines based on the feature engineering requirements, set up automated data quality checks, and ensure that the data flow supports the prototyping pace.

Nandani: Optimize cloud services for model prototyping, ensuring that computational resources are efficiently allocated and scaled according to the needs of the ML team.

Mahrukh: Work with Arham to interpret the features' effects on the models, and correlate these with industry practices and potential user impact, updating the market analysis with new data insights.

Phase 4: Model Refinement and Experimentation

Aasna: Facilitate sprint planning sessions to prioritize tasks for model refinement and coordinate interdepartmental efforts to support the experimentation phase.

Arham: Perform extensive hyperparameter tuning on the bestperforming prototype models, implement crossvalidation techniques, and work closely with Krishan to analyze experiment results.

Krishan: Enhance the ML pipeline with automated model training and evaluation processes, and ensure robust logging for all experiments for review and audit purposes.

Yash: Ensure the data pipeline can handle iterative model training sessions, including realtime data updates and automated feeding into the model training queue.

Nandani: Implement autoscaling for computational resources based on the demand from model training sessions and set up monitoring for cloud performance and cost management.

Mahrukh: Analyze interim model performance reports and provide recommendations on potential business implications, adjusting the market strategy as needed.

Phase 5: Deployment Preparation and Testing

Aasna: Coordinate a predeployment review to ensure all product components meet quality standards, oversee the preparation of deployment documentation, and ensure alignment with the gotomarket strategy.

Arham: Collaborate with Krishan to containerize the final model, prepare deployment scripts, and conduct a final round of validation tests.

Krishan: Set up a CI/CD pipeline for the deployment of the machine learning model, integrate monitoring tools for postdeployment model performance, and prepare rollback procedures.

Yash: Validate the production data pipeline for consistency and performance, and ensure that it is robust and secure for deployment.

Nandani: Establish a productionready cloud environment, complete final load and security testing, and confirm disaster recovery and data backup procedures are in place.

Mahrukh: Finalize the product impact analysis, prepare marketing and communication material based on product capabilities and expected outcomes, and document the user acceptance testing plan.

Phase 6: Model Deployment and Monitoring

Aasna: Oversee the model deployment to production, coordinate the monitoring strategy across teams, and manage any immediate postdeployment issues.

Arham: Monitor the model's realworld performance, analyze discrepancies from expected outcomes, and prepare for iterative model updates based on realworld feedback.

Krishan: Implement continuous monitoring for model performance, setting up alerts for drift detection, and ensuring model servicing and updating processes are in place.

Yash: Ensure realtime data feeds into the production model without interruption, and maintain data pipeline health, including monitoring for anomalies or disruptions.

Nandani: Monitor the cloud infrastructure for uptime, performance, and costeffectiveness, and apply optimizations as necessary.

Mahrukh: Continuously analyze market feedback and user data to refine the customer profile and value proposition, and prepare reports for stakeholders on the product's market performance.